Supporting Information

End-to-end Attraction of Duplex DNA

Christopher Maffeo, Binquan Luan, and Aleksei Aksimentiev

August 26, 2011

Here we describe the details of the simulations presented in the main text following the order of Results and Discussion.

1 Mechanics of end-to-end dissociation

1.1 Constant force simulations

Several simulations were performed in which a constant force was applied along the axis of the DNA helix. The simulations used the same initial conditions as the SMD simulations described below in Sec. 1.3. The phosphorous atoms of one DNA fragment were harmonically restrained ($k_s = 445$ pN/nm) about their initial coordinates, while a force (100, 150, or 200 pN) was applied to the phosphorous atoms of the second DNA fragment along the axis of the first DNA fragment. A 100 or 150 pN force did not rupture the assembly (simulations performed for 30 ns and 50 ns, respectively). By contrast, rupture occurred within 5 ns with a force of 200 pN. We note that tilting of the terminal base pairs, which we associate with shearing of the end-to-end assembly, was suppressed by applied restraints; the force needed to rupture the assembly would likely have been lower in the absence of these restraints.

1.2 Axial stretching of the end-to-end assembly

Force induced dissociation of the end-to-end DNA assembly was simulated using the steered molecular dynamics (SMD) method [1]. Figure S1 a schematically illustrates the process. Each DNA fragment was attached to one end of a virtual spring. The ends of the spring were anchored to the center of mass (CoM) of either the phosphorous atoms of the DNA fragments (6 simulations) or the CoM of the closest DNA ends (4 simulations). The rest length of the spring was increased at a constant rate of 0.4 or 0.2 Å/ns until the rupture of the assembly occurred. The work performed during each of these simulations was recorded and combined using the Jarzynski's equality [2] to estimate the potential of mean force (PMF) for the process considered:

$$e^{-\beta\,\Delta F} = \left\langle e^{-\beta\,W} \right\rangle$$

The angle brackets denote an ensemble average; β is $1/\mathrm{k_B T}$; ΔF is the change in free energy when the system is brought from one state to another; W is the work done during the change of state. The work performed was found to be independent of the pulling rate.

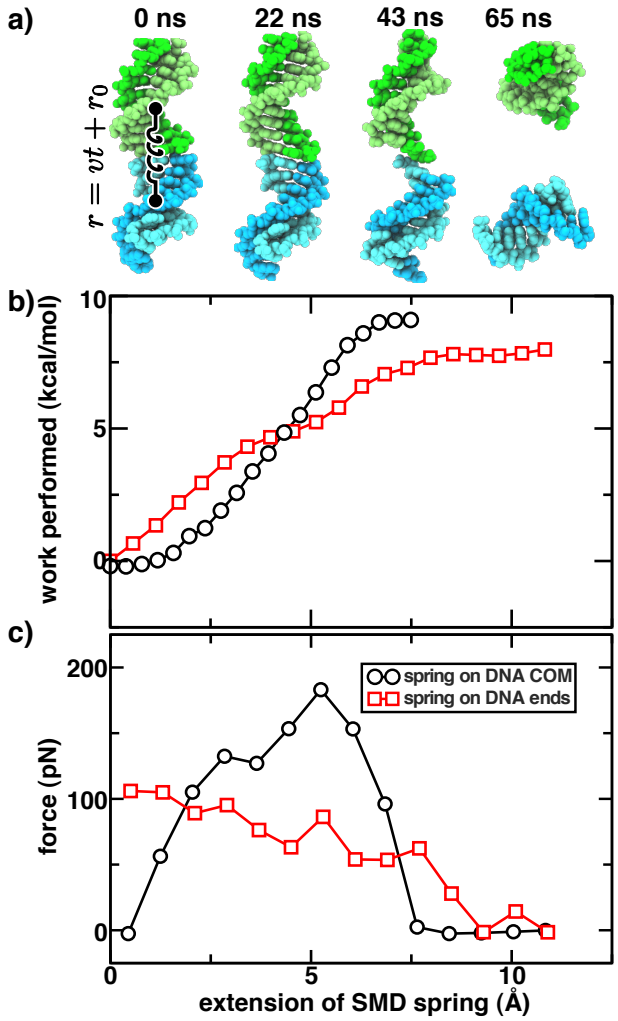

Figure S1: Axial stretching of the end-to-end DNA assembly. (a) Snapshots illustrating the conformations of DNA fragments in a typical simulation of axial stretching performed using a spring tethered to the CoM of the fragments. The DNA fragments are drawn using green and blue van der Waals spheres. (b) Average work performed during the stretching simulations. The force of the virtual spring was applied either to the centers of mass (black circles; 6 simulations) or to the ends (red squares; 4 simulations) of the DNA fragments. (c) Force applied by the spring during the axial stretching simulations, averaged in blocks (roughly 1 block/Å extension) using data from multiple simulation trajectories.

The free energy difference between bound and unbound states was $\sim$ 8.5 kcal/mol regardless of the choice for the spring anchors, Fig. S1 b. However, the variation of the applied force with the rest length of the spring (Fig. S1 c) depended strongly on the simulation protocol employed, reflecting differences in the rupture pathway. Figure S1 a depicts the rupture process of the end-to-end assembly by a spring tethered to the CoM of the DNA: the DNA fragments stretch until the terminal base pairs tilt and the ends separate by shearing. When the spring force was applied to the ends of the fragments, the rupture occurred by shearing but without significant deformation of the DNA.

1.3 Transverse shearing of the end-to-end assembly

In the axial stretching simulations, the rupture of the end-to-end assembly occurred by shearing of the terminal base pairs regardless of where the spring was anchored. We carried out a set of SMD simulations

designed to produce shearing deformation: pulling one DNA fragment with respect to the other in a direction perpendicular to the common axis of the DNA assembly. Prior to production simulation, an end-to-end complex of two DNA fragments was equilibrated in the presence of restraints that aligned its symmetry axis with the z axis of the simulation cell. In the resulting configuration, the major groove of the end-to-end junction nearly faced the y axis of our coordinate system, while the minor groove nearly faced the $-y$ axis. During production simulations, one end of a spring ($k_s = 4000$ pN/nm) was tethered to the terminal base pair of one DNA fragment while the other end of the spring was pulled in the $-x$, $-y$, $+x$, or $+y$ direction at a rate of 0.2 Å/ns. All phosphorous atoms of the other DNA fragment were harmonically restrained ($k_s = 13,900$ pN/nm per atom) about their initial coordinates.

Figure S2 a depicts a typical MD trajectory resulting from the SMD simulations of transverse shearing. The work performed during each of the four simulations is shown in Fig. S2 b. By the time of rupture, which is indicated by an open circle, the work reached ∼10 kcal/mol. These simulations, which did not mechanically deform the DNA, confirm the surprisingly high energy required to rupture the assembly during the SMD simulations of axial stretching.

1.4 Rupture of end-to-end assemblies containing complementary ssDNA overhangs

Five systems containing 20 DNA base pairs were built to determine the effect of complementary overhangs on the stability of the end-to-end DNA complex. The first system was a continuous 20-base pair (two-turn) helical dsDNA fragment immersed in 0.1 M aqueous solution of NaCl. The other four systems were built from the first system by cleaving the dsDNA into a pair of equal-length 5′-phosphorylated dsDNA fragments containing 0-, 1-, 2- or 4-nucleotide, complementary ssDNA tails, Fig. S3 a. Following usual minimization and equilibration protocols, the DNA assembly was axially stretched using the SMD method described in Sec. 1.2. The force of the virtual spring was applied to the CoM of the DNA fragments.

Figure S3 b shows the average work performed, estimated for each system from two independent SMD trajectories. The average work profile in the absence of ssDNA overhangs (black circles) agrees well with the result of our previous simulation of a similar system (grey circles), which began with a different conformation of the DNA fragments. The work profiles for stretching two blunt-ended DNA fragments begins to diverge from that of a continuous dsDNA fragment (closed left-facing blue triangles) as the base pairs forming the junction in the former begin to shear. Rupturing the end-to-end complex of two DNA fragments having 1-nucleotide overhangs (red squares) requires ∼ 4.5 kcal/mol more work than rupturing blunt-ended fragments; the presence of ssDNA overhangs suppresses the shearing of base pairs at the end-to-end junction at low forces. In the case of the fragments having 2- or 4-nucleotide overhangs, the average work profiles (magenta diamonds and purple triangles) initially follows that of a continuous dsDNA but diverges from it at larger extensions once base stacking in the end-to-end assemblies ruptures, leaving stretched ssDNA tails that maintain contact over the end-to-end junction.

From these simulations, we conclude that the presence of sticky overhangs makes the end-to-end assembly similar to a continuous duplex DNA at load forces smaller than that required to overstretch a dsDNA helix. In comparison to a complex of two blunt-ended DNA fragments, rupture of the fragments having ssDNA overhangs requires additional work, which, in the case of single nucleotide overhangs, we found to be in good agreement with the free energy required to rupture an isolated AT base [3].

2 Potential of mean force of axially aligned DNA duplexes

Umbrella sampling simulations were used to obtain the relative free energy differences for end-to-end association between systems with various ion concentrations, chemical structures of DNA termini, and relative azimuthal angles. The main findings of these simulations are reported in the results section of the main text, and the protocols used to assemble the systems and perform the simulations are detailed in the methods section of the main text. Here we provide results that complement Table 1 in the main text, additional discussion, and descriptions of additional simulations.

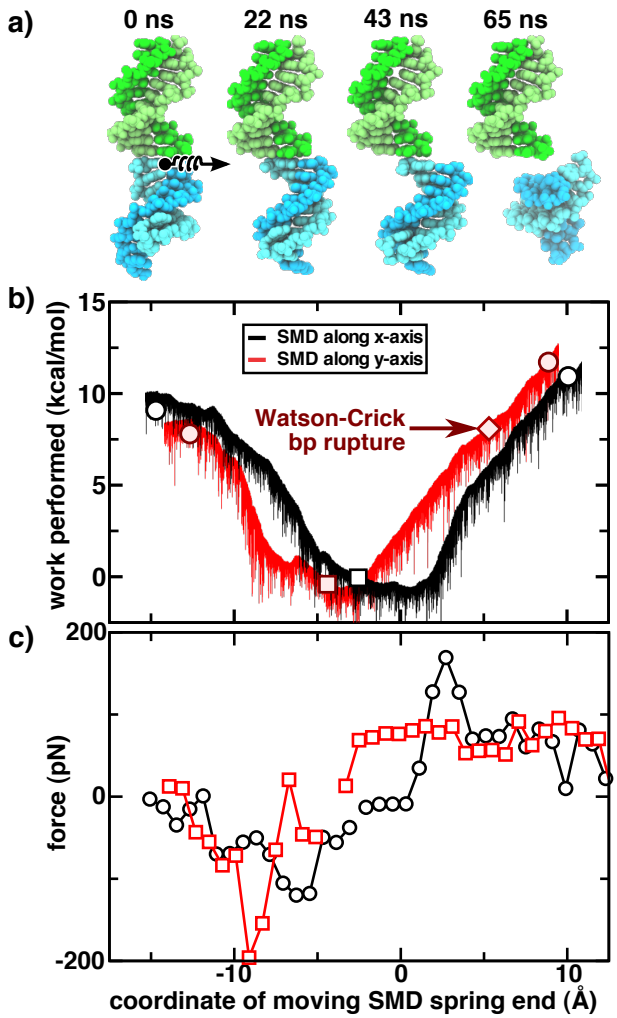

Figure S2: Transverse shearing of the end-to-end assembly. (a) The rupture process. The phosphorous atoms of one DNA fragment (top; green) were harmonically restrained about their initial coordinates while the other DNA fragment (bottom; blue) was pulled along the $-x$ axis. The DNA atoms are depicted as in Fig. S1 a. (b) Work required to shear the DNA ends. Plotted is the work performed by the SMD spring as a function of the coordinate of the moving end of the spring along the x (black) or y (red) axis. The spring force was applied to the CoM of the terminal base pair of the moving fragment. The symbols denote the state of the system at the beginning of the simulations (squares) and at the moment of rupture (circles). In one simulation, the terminal base pair tethered to the SMD spring ruptured; this moment is indicated by a diamond. (c) The force of the spring applied to the DNA fragment, averaged in ∼1-Å-extension blocks.

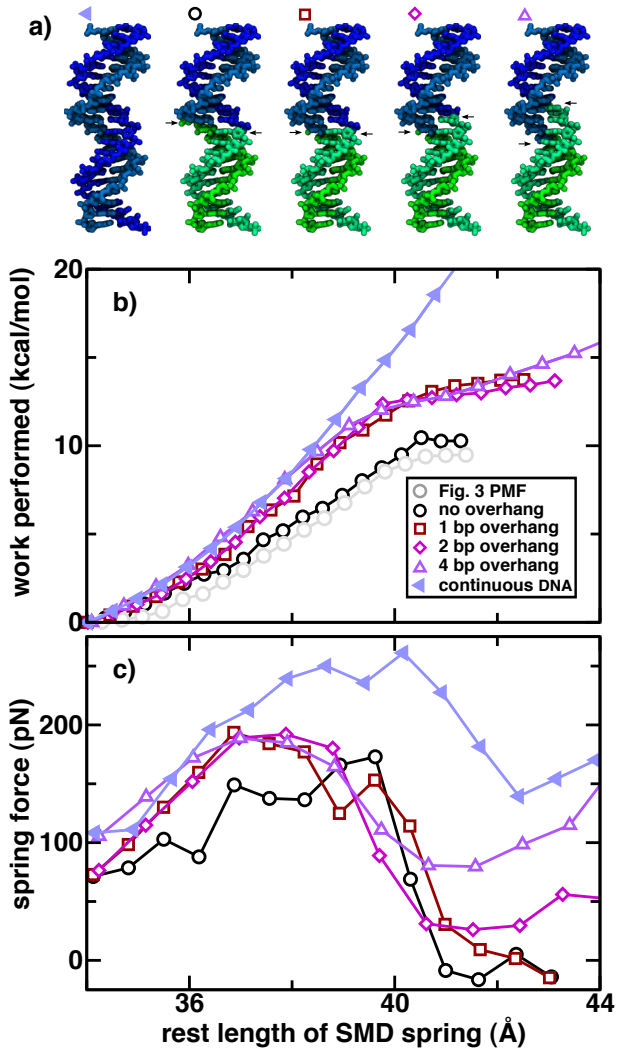

Figure S3: Axial stretching of end-to-end assemblies having complementary ssDNA overhangs. (a) Five DNA systems were explored in this set of simulations. In addition to a continuous 20-base pair dsDNA helix (closed left-facing blue triangles), the systems contained two disjoint DNA fragments with 0 (black circles), 1 (red squares), 2 (magenta diamonds), or 4 (purple triangles) complementary ssDNA overhangs. The arrows indicate the positions of nicks in the DNA assemblies. (b) The average work performed by the spring during axial stretching. For comparison, grey circles reproduce the results of axial stretching shown in Fig. S1 b (grey circles). (c) The force of the spring averaged in 1-Å blocks of spring extension.

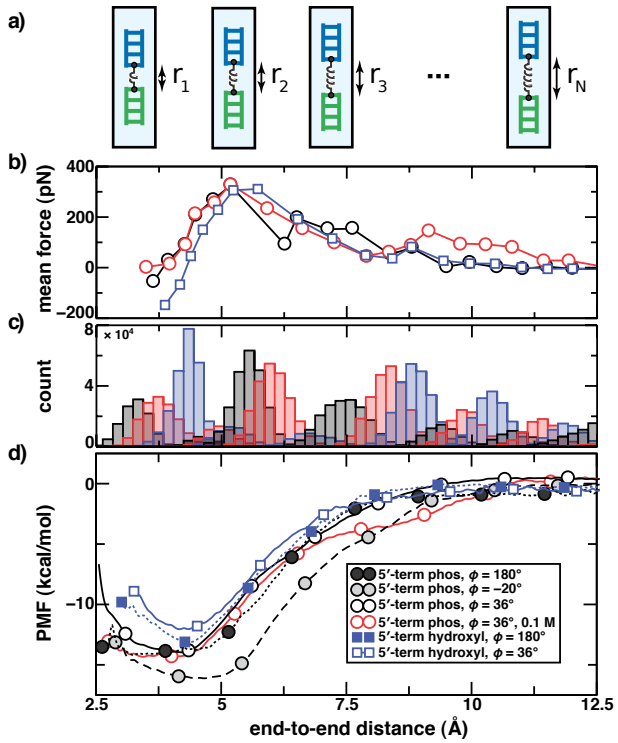

Figure S4: Umbrella sampling simulations of aligned dsDNA fragments. (a) Schematics illustrating the method used to obtain the force and the PMF. In addition to the umbrella sampling potential, illustrated by a spring, harmonic restraints maintained the axial alignment of the DNA fragments. (b) The mean force exerted by the umbrella sampling spring on 5′-terminal phosphates DNA fragments in 1.0 M (black circles) and 0.1 M (red circles) solution of NaCl and blunt-ended DNA fragments in 1.0-M (blue squares) NaCl. The relative azimuthal angle ϕ was restrained about 36°. The force applied by the umbrella sampling potential was averaged over a simulation trajectory to produce each data point. (c) Distributions of the end-to-end distances measured in the ensemble containing 5′-phosphorylated DNA in 1.0 M electrolyte with ϕ restrained about 36°. (d) PMF of the end-to-end interaction, extracted from distributions like those plotted in (c). All simulations were performed in 1.0-M solution of NaCl, except those yielding the data depicted with red circles in (b) and (d), which were performed in 0.1 M solution of NaCl.

Figure S4 a schematically illustrates the umbrella sampling protocol used to obtain the relative free energies reported in Table 1 of the main text. An ensemble of simulations was performed on a two-fragment DNA system using a harmonic spring to maintain a prescribed end-to-end separation, which differed between the simulations. The average force applied by the spring in several of the ensembles is plotted as a function of the average end-to-end separation in Fig. S4 b. In these particular simulations, the force is nearly zero for distances above 10 Å, and rapidly increases to a peak value of ∼300 pN at 5.5 Å end-to-end separation. Fig. S4 c demonstrates that such simulations generate overlapping distributions of the end-to-end separation. Obtained thereby, distributions for each ensemble were analyzed using the weighted histogram analysis method (WHAM) to obtain the PMFs plotted in Fig. S4 d. The strength of the interaction obtained here (11–17 kcal/mol) is considerably larger than suggested by the PMF in Fig. 3 of the main text (∼ 7 kcal/mol). We attribute the discrepancy primarily to the stiffer axial restraining potentials used during the above simulations (139 vs. 13.9 pN/nm per atom) as well as to the azimuthal restraining potential, which was not applied in the simulations reported in Fig. 3 of the main text. Since the axial and azimuthal restraints involved the same spring constants in all simulations reported in Fig. S4, their effect on the PMF is most likely a static offset insensitive to the variations between the ensembles. The relative free energies for end-to-end association between different ensembles remain physically meaningful although the restraints distort the absolute free energy values.

The azimuthal restraints were enforced by a torque whose magnitude and direction depended harmonically on ϕ. The torque was distributed among the phosphorous atoms of the DNA rather than the atoms whose coordinates determined the instantaneous value of ϕ. Consequently the azimuthal restraints did not constitute a proper potential. In general the use of restraints that are not proper potentials should be avoided during umbrella sampling simulations because the free energy associated with the application of such restraints is poorly defined.

3 Standard binding free energy of end-to-end association

3.1 Calculations of G_{bind}

The axial restraints used in previous umbrella sampling simulations enhanced the free energy of end-to-end association by reducing the conformational space available to separated DNA fragments. Because the axial restraints depended on coordinates that are not orthogonal to the end-to-end distance, we are formally unable to estimate the impact of the restraints on the free energy. The standard binding free energy, G_{bind}, for the end-to-end assembly was obtained using a more general, multi-step process roughly following that presented by Roux and coworkers that properly accounts for the confinement of the axial alignment potential. Here we describe the details and possible limitations of each step of the employed method.

The equilibrium binding constant, K_{eq}, which is related to G_{bind}, can be written in terms of the fraction of time a pair of molecules form a bound complex. In the limit of a dilute DNA solution, this can be written as a ratio of integrals over the Boltzmann-weighted configurational space in which the DNA fragments are bound and unbound:

$$K_{\mathrm{eq}} = \frac{1}{2}\frac{4\int_b d\mathbf{X}\;\; e^{-\beta U}}{\int d\mathbf{X}\;\delta\left(\mathbf{r}-\mathbf{r}^*\right)e^{-\beta U}} = \frac{1}{C_0}e^{\beta G_{\mathrm{bind}}}\;.$$

Here, $\mathbf{X}$ represents the coordinates of all atoms in our system of volume $V = 1/(2c_{\mathrm{DNA}})$, where c_{DNA} is the effective concentration of DNA fragments, $U = U[\mathbf{X}]$ is the Hamiltonian, $\mathbf{r}$ represents the vector connecting the terminal base pairs of the DNA, and $\mathbf{r}^*$ is some reference separation sufficiently large that the DNA fragments no longer interact. The subscript b denotes that the configurational integral is restricted to conformations in which the DNA fragments are bound. The delta function in the denominator was used to absorb a factor of V in the numerator. The factor of one half before the ratio is due to the stoichiometry of homo-dimerization. The factor of four in the numerator is due to symmetry; we consider b for only one of four unique ways of pairing the DNA ends. Finally, C_0 is the standard state DNA concentration of 1 M.

The expression for K_{eq} can be written as an integral over the PMF. Computational expense requires additional restraining potentials, U_θ, that limit the amount of phase space explored by the DNA fragments

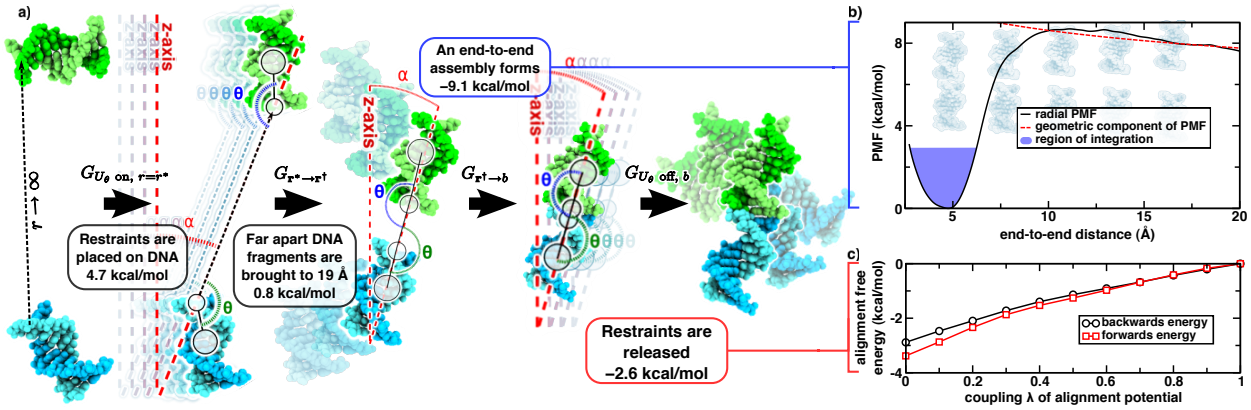

Figure S5: Calculation of the standard binding free energy of the end-to-end DNA complex. (a) The standard binding free energy was obtained in four steps. The DNA atoms are depicted as blue and green spheres. (b) The PMF, $W(r)$, (solid black line) in the presence of U_θ. When $r > 10$, the interaction between the DNA fragments is negligible and $W(r)$ equals $-k_BT \ln r^2$ (red curve) due to the entropic push towards larger r. The blue shaded area indicates the integration region used to obtain $G_{\mathbf{r}^\dagger\to b}$. (c) The free energy change of reducing the fully coupled potential U_θ to $\lambda_i U_\theta$ as obtained by the evaluation of the forwards (black) and backwards (red) FEP averages (see text) with $\lambda_i = \frac{10-i}{10}$ for $i = [0, 10]$.

(e.g. axial alignment potentials). However, such potentials impact the PMF and need to be accounted for by multiplying the expression for K_{eq} by appropriate factors that effectively break G_{bind} into steps that represent free energy changes due to turning the potentials on, bringing the fragments together, and turning the potentials off.

$$G_{\text{bind}} = G_{U_\theta \text{ on, } r=r^*} + G_{\mathbf{r}^*\to\mathbf{r}^\dagger} + G_{\mathbf{r}^\dagger\to b} + G_{U_\theta \text{ off, } b} \,.$$

To be computationally efficient, a small simulation system was used, so we introduced the free energy $G_{\mathbf{r}^*\to\mathbf{r}^\dagger}$ required to bring the DNA fragments from infinite separation to the furthest separation in our umbrella sampling simulations. We have,

$$G_{U_\theta \text{ on, } \mathbf{r}=\mathbf{r}^*} = -k_BT \ln \left[\frac{\int d\mathbf{X}\, \delta\left(\mathbf{r}-\mathbf{r}^*\right) e^{-\beta(U+U_\theta)}}{\int d\mathbf{X}\, \delta\left(\mathbf{r}-\mathbf{r}^*\right) e^{-\beta U}} \right]$$

$$G_{\mathbf{r}^*\to\mathbf{r}^\dagger} = -k_BT \ln \left[\frac{\int d\mathbf{X}\, \delta\left(\mathbf{r}-\mathbf{r}^\dagger\right) e^{-\beta(U+U_\theta)}}{\int d\mathbf{X}\, \delta\left(\mathbf{r}-\mathbf{r}^*\right) e^{-\beta(U+U_\theta)}} \right]$$

$$G_{\mathbf{r}^\dagger\to b} = -k_BT \ln \left[2C_0 \frac{\int_b d\mathbf{X}\, e^{-\beta(U+U_\theta)}}{\int d\mathbf{X}\, \delta\left(\mathbf{r}-\mathbf{r}^\dagger\right) e^{-\beta(U+U_\theta)}} \right]$$

$$G_{U_\theta \text{ off, } b} = -k_BT \ln \left[\frac{\int_b d\mathbf{X}\, e^{-\beta U}}{\int_b d\mathbf{X}\, e^{-\beta(U+U_\theta)}} \right]$$

U_θ must be carefully chosen to ensure that each step can be readily evaluated; here we have selected $U_\theta = \frac{k_\theta}{2}((\theta_1 - 180°)^2 + (\theta_2 - 180°)^2)$ to guide the DNA into a bound complex, where $k_\theta = 14$ kcal/mol rad^2 ($\sim$12° RMSF) and θ_i is the angle between the CoM of DNA molecule i, its terminal base pair, and the terminal base pair of the other DNA molecule. The free energy decomposition is illustrated in Fig. S5 a.

Evaluation of $G_{U_\theta \text{on, } r=r^*}$

Since the DNA fragments are not interacting when $\mathbf{r} = \mathbf{r}^*$, almost all degrees of freedom are identical in the numerator and denominator of the expression for $G_{U_\theta \text{ on, } r=r^*}$, allowing direct integration to obtain this energy exactly

$$G_{U_\theta \text{ on, } r=r^*} = -k_B T \ln \left[\frac{\int_0^\pi d\theta \, \sin\theta \, e^{-\beta \frac{k_\theta}{2} \theta^2}}{\int_0^\pi d\theta \, \sin\theta} \right]^2 = 4.72 \text{ kcal/mol}$$

Evaluation of $G_{\mathbf{r}^* \to \mathbf{r}^\dagger}$

Under the Debye-Hückel approximation, interactions between point particles are modeled through the Yukawa potential, $U_{\text{Yukawa}}(r) = \frac{q_1 q_2}{4\pi\epsilon\epsilon_0} \frac{e^{-r/\lambda_D}}{r}$, where λ_D is the ion concentration-dependent Debye length. When $\mathbf{r} = \mathbf{r}^*$, we take this energy to be zero. When $\mathbf{r} = \mathbf{r}^\dagger$, we assume the interaction between the DNA fragments can be modeled by pair Yukawa potentials. The Hamiltonian becomes $U = U_0 + U_{\text{DH}}$, where U_0 is the Hamiltonian governing the internal conformation of the DNA fragments and U_{DH} is the sum of U_{Yukawa} over all pairs of atoms composed from both DNA fragments. Since $e^{-\beta(U_0+U_\theta)}$ is independent of the value of $\mathbf{r}$,

$$\begin{aligned}
G_{\mathbf{r}^* \to \mathbf{r}^\dagger} &= -k_B T \ln \left[\frac{\int d\mathbf{X} \, \delta\left(\mathbf{r} - \mathbf{r}^\dagger\right) e^{-\beta(U_{\text{DH}}+U_0+U_\theta)}}{\int d\mathbf{X} \, \delta\left(\mathbf{r} - \mathbf{r}^*\right) e^{-\beta(U_0+U_\theta)}} \right] \\
&= -k_B T \ln \left[\frac{\int d\mathbf{X} \, \delta\left(\mathbf{r} - \mathbf{r}^\dagger\right) e^{-\beta(U_{\text{DH}}+U_0+U_\theta)}}{\int d\mathbf{X} \, \delta\left(\mathbf{r} - \mathbf{r}^\dagger\right) e^{-\beta(U_0+U_\theta)}} \right] \\
&= k_B T \ln \left[\frac{\int d\mathbf{X} \, \delta\left(\mathbf{r} - \mathbf{r}^\dagger\right) e^{\beta U_{\text{DH}}} e^{-\beta(U_{\text{DH}}+U_0+U_\theta)}}{\int d\mathbf{X} \, \delta\left(\mathbf{r} - \mathbf{r}^\dagger\right) e^{-\beta(U_{\text{DH}}+U_0+U_\theta)}} \right] \\
&= k_B T \ln \langle e^{\beta U_{\text{DH}}} \rangle_{\mathbf{r}=\mathbf{r}^\dagger} = 0.75 \pm 0.1 \text{ kcal/mol}
\end{aligned}$$

We estimated the value of $\langle e^{\beta U_{\text{DH}}} \rangle_{\mathbf{r}=\mathbf{r}^\dagger}$ by averaging over DNA conformations obtained from a simulation with r restrained about 19 Å in 0.1 M electrolyte. The free energy required to bring the ends of two DNA fragments to 19 Å separation is negligible in 1 M electrolyte.

Evaluation of $G_{\mathbf{r}^\dagger \to b}$

$G_{\mathbf{r}^\dagger \to b}$ can be written in terms of an integral of $W(r')$ over the bound volume, where $W(r') = -k_B T \ln \frac{\int d\mathbf{X} \, \delta\left(r - r'\right) e^{-\beta(U+U_\theta)}}{\int d\mathbf{X} \, e^{-\beta(U+U_\theta)}}$ is the (1-dimensional) PMF [4, 5].

$$\begin{aligned}
G_{\mathbf{r}^\dagger \to b} &= -k_B T \ln \left[2C_0 \frac{\int_b d\mathbf{X} \, e^{-\beta(U+U_\theta)}}{\int d\mathbf{X} \, \delta\left(\mathbf{r} - \mathbf{r}^\dagger\right) e^{-\beta(U+U_\theta)}} \right] \\
&= -k_B T \ln \left[2C_0 \frac{\int_b d\mathbf{X} \int dr' \, \delta\left(r - r'\right) e^{-\beta(U+U_\theta)}}{\int d\mathbf{X} \, \frac{1}{4\pi {r^\dagger}^2} \delta\left(r - r^\dagger\right) e^{-\beta(U+U_\theta)}} \right] \\
&= -k_B T \ln \left[2C_0 \frac{\int_b dr' \int d\mathbf{X} \, \delta\left(r - r'\right) e^{-\beta(U+U_\theta)}}{\frac{1}{4\pi {r^\dagger}^2} \int d\mathbf{X} \, \delta\left(r - r^\dagger\right) e^{-\beta(U+U_\theta)}} \right] \\
&= -k_B T \ln \left[2C_0 \, 4\pi {r^\dagger}^2 \frac{\int_b dr' \, e^{-\beta W(r')}}{e^{-\beta W(r^\dagger)}} \right] \\
&= -k_B T \ln \left[2C_0 \int_b dr' \, 4\pi {r^\dagger}^2 e^{-\beta(W(r') - W(r^\dagger))} \right]
\end{aligned}$$

Note that the above derivation requires that U_θ is independent of $\mathbf{r}$, and that it was assumed that the value of r alone can describe whether the DNA fragments are in a bound conformation.

The umbrella sampling method was employed to obtain $W(r)$ using an ensemble of anisotropic systems (see Fig. 1 a of the main text). Simulations were performed with the end-to-end distance r restrained about values from 3.5–12 Å in 0.5 Å intervals with a spring constant of 4000 pN/nm, and about values from 13–19 Å in 1 Å intervals with a spring constant of 1000 pN/nm. Unlike the umbrella sampling simulations described in the main text, r was not projected along the common DNA axis. The ensemble was initially created by placing the DNA fragments in an anisotropic system so that the spring was at its rest length. To mitigate bias due to the initial DNA conformation, four systems were created for each restraint distance with the relative azimuthal angle $\phi(0) = 0$, 90, 180, and 270°. In addition to the umbrella sampling potential, U_θ was applied to maintain axial alignment of the DNA, and a harmonic potential restrained the angle between $\mathbf{r}$ and the z axis (α in Fig. S5) about 0°. The latter potential has no effect on $W(r)$, and does not affect the configurational integrals. The water and ions in each simulation system were equilibrated for 10 ps while stiff springs (initially 1 kcal/mol Å^2) that restrained each non-hydrogen DNA atom to its initial position were removed. After equilibrating for 4 ns, data was accumulated for another 4 ns.

$W(r)$ represents the interactions between a pair of DNA fragments in isolation. However, in our simulations the DNA fragments interact with periodic images of one another. We attempted to remove effects due to the proximity of periodic images by assuming that our Hamiltonian is composed of interactions between the pair of DNA fragments in the same unit cell (including U_θ), U_{cell}, as well as interactions between one DNA fragment and the nearest periodic image of the other DNA fragment, U_{per}, which we do not wish to include in $W(r)$ or G_{bind}. We model U_{per} as a sum of the Yukawa potential between all pairs of atoms composed from a DNA fragment and the nearest periodic image of the other DNA fragment along the common DNA axis.

We obtained $W(r') - W(r^\dagger)$ by solving for the average value of $e^{\beta U_{\text{per}}}$ as a function of r' using the WHAM equations [4], since

$$W(r') - W(r^\dagger) = -\,k_B T \ln \frac{\int d\mathbf{X}\; \delta\left(r - r'\right) e^{\beta U_{\text{per}}} e^{-\beta(U_{\text{cell}} + U_{\text{per}})}}{\int d\mathbf{X}\; \delta\left(r - r^\dagger\right) e^{\beta U_{\text{per}}} e^{-\beta(U_{\text{cell}} + U_{\text{per}})}}$$

Removing interactions over the periodic boundary in this way made the minimum of $W(r) - W(r^\dagger) \sim 0.2$ kcal/mol deeper.

$G_{\mathbf{r}^\dagger \to b}$ was obtained by integrating $4\pi r^{*2} e^{-\beta(W(r) - W(r^*))}$ over values of r for which the DNA fragments are bound, taken to be the values of r for which $W(r) - \min(W) < 5\ k_B T$ (blue region indicated in Fig. S5 b). Thus, we obtained $G_{\mathbf{r}^\dagger \to b} = -9.1$ kcal/mol. We note that integration over slightly different regions yield almost identical results provided the lowest several $k_B T$ of $W(r)$ are included.

Evaluation of $G_{U_\theta \text{off},\, b}$

Finally, the free energy gained upon releasing the axial restraints $G_{U_\theta\ \text{off},\ b}$ was obtained using the free energy perturbation (FEP) technique [6, 7] to remove U_θ in 10 steps, Fig. S5 c. When the DNA fragments are bound, the values of θ_i are coupled to one another and to other degrees of freedom in $\mathbf{X}$, and the integrals cannot be computed directly. However, for Hamiltonians U_1 and U_2 each acting on $\mathbf{X}$ degrees of freedom,

$$\begin{aligned}
\frac{\int d\mathbf{X}\; e^{-\beta U_1}}{\int d\mathbf{X}\; e^{-\beta U_2}} &= \frac{\int d\mathbf{X}\; e^{-\beta(U_1 - U_2)} e^{-\beta U_2}}{\int d\mathbf{X}\; e^{-\beta U_2}} \\
&= \left\langle e^{-\beta(U_1 - U_2)} \right\rangle_{U_2}, \text{or equivalently} \\
&= \left\langle e^{\beta(U_1 - U_2)} \right\rangle_{U_1}^{-1},
\end{aligned}$$

where the angle brackets represent a thermodynamic average in the ensemble indicated by the subscript. Thus, we have

$$G_{U_\theta\ \text{off},\ b} = -k_B T \ln \left\langle e^{-\beta U_\theta} \right\rangle_U^{-1} = -k_B T \ln \left\langle e^{\beta U_\theta} \right\rangle_{U + U_\theta}$$

In the limit of excellent sampling, either average can be obtained during an MD simulation. However, if U_θ represents a significant change of the Hamiltonian, rarely sampled regions of phase space dominate the estimates of these averages. To circumvent this issue, U_θ was removed during 10 relatively small steps that each perturbed system only weakly, in accordance with FEP [6, 7]. The coefficients $\lambda_i = \frac{10-i}{10}$ with $i = [0, 10]$ coupled U_θ to the system,

$$\begin{aligned} G_{U_\theta \text{ off, } b} &= -k_B T \ln \left[\frac{\int_b d\mathbf{X}\ e^{-\beta(U+\lambda_{10}U_\theta)}}{\int_b d\mathbf{X}\ e^{-\beta(U+\lambda_9 U_\theta)}} ... \frac{\int_b d\mathbf{X}\ e^{-\beta(U+\lambda_1 U_\theta)}}{\int_b d\mathbf{X}\ e^{-\beta(U+\lambda_0 U_\theta)}} \right] \\ &= -k_B T \sum_i \ln \left\langle e^{-\beta(\lambda_{i+1}-\lambda_i)U_\theta} \right\rangle_{U+\lambda_i U_\theta} \\ &= -k_B T \sum_i \ln \left\langle e^{\beta(\lambda_{i+1}-\lambda_i)U_\theta} \right\rangle^{-1}_{U+\lambda_{i+1} U_\theta} \end{aligned}$$

The FEP simulation systems were prepared from the same systems used to obtain $W(r)$, with r restrained about 3.5 Å. The systems were each equilibrated for 10 ns without restraining r or θ_i. For each of the four initial relative azimuthal angles, another four simulations were created with 2.5 ns of additional equilibration between each of these. Thus, 16 production simulations began with $\lambda = \lambda_{10} = 0$ and lasted 0.4 ns each. When final coordinates were used to begin a simulation using the next value of λ, until 16 simulations had been performed for each successive value of λ_i. The values of θ_i were recorded every 40 fs. The free energy change of bringing $\lambda_i \rightarrow \lambda_{i+1}$ was evaluated directly using $\langle e^{-\beta(\lambda_{i+1}-\lambda_i)U_\theta} \rangle_i^{-1}$ (forwards direction; black in Fig. S5 c) and, $\langle e^{\beta(\lambda_{i+1}-\lambda_i)U_\theta} \rangle_{i+1}$ (backwards direction; red in Fig. S5 c). The discrepancy between these values provides a measure of the statistical error of our simulations and would be zero in the limit of perfect sampling. The sum of these free energy differences provides $G_{U_\theta \text{ off, } b} = -2.56 \pm 0.4$ kcal/mol Since the DNA in the end-to-end complex are approximately axially aligned, removing U_θ only partially recovered the free energy cost of applying U_θ.

3.2 Estimate of rate of end-to-end dissociation

The escape rate for a Brownian particle trapped in a potential well can be estimated by finding the mean first passage time, τ, from the Smoluchowski equation, which describes the motion of a Brownian particle in a potential. It can be shown that for a particle with a 1-dimensional diffusion coefficient D [8],

$$\tau = \frac{1}{D} \int_{r_{\min}}^{r_{\max}} dy\ e^{W(y)/k_B T} \int_0^y dz\ e^{-W(z)/k_B T},$$

where $r_{\min}$ is the minimum of the PMF, $W(r)$, and $r_{\max}$ is the height of the barrier. Since $r_{\max}$ is an inflection point, the rate of crossing, k_{off}, is 1/2 the rate of arrival $1/\tau$. This approach was used previously to obtain k_{off} from D and W, which in turn were obtained through MD simulations [9].

The value of k_{off} is sensitive to the location of the peak PMF, $r_{\max}$. Estimating that $D \sim 25$ Å^2/ns from trajectories of the end-to-end distance obtained in **Collapse of aligned dsDNA** of the main text, the range of reasonable choices for the location of the maximum (10–15 Å) gives $k_{\text{off}}^{-1} \sim$ 170,000–860,000 ns. We estimate that $r_{\max}$ is at 12 Å, yielding $k_{\text{off}}^{-1} \sim$ 480,000 ns. However, the use of axial alignment restraints in our calculations of the PMF may have affected the above estimates of k_{off} by reducing the pathways available to the DNA fragments for escape. Thus, we expect the above numbers represent an upper bound estimate for k_{off}^{-1}.

4 Spontaneous assembly of long end-to-end aggregates

458 DNA fragments 10 bp in length were placed at in a cube of NaCl solution 23.8 nm on each side (after NPT equilibration) making the system shown in Fig. 4 a of the main text with a DNA concentration of 56.4 mM (349 mg/ml). The system was neutralized prior to the addition of 0.1 mM NaCl, but the Donnan

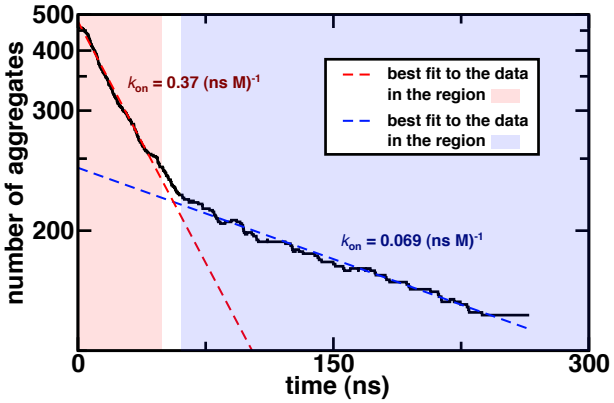

Figure S6: The number of end-to-end aggregates (including monomers) during a simulation starting with 458 DNA monomers in a cube of solvent. The reversible step-growth polymerization model predicts 2$^{\text{nd}}$ order kinetics for the total number of aggregates. We plot the best fit of this data in two regions: during the first 49.5 ns (red) and after 60.9 ns (blue). The fits provided estimates of k_{on} for the end-to-end association of two polymers.

equilibrium approximation for charged cylinders suggests that this represents coupling the system to a bath containing 350 mM NaCl [10, 11]. During a 260-ns MD simulation, the DNA fragments diffused about their initial positions and interacted with their neighbors to form aggregates up to 11 DNA fragments (110 base pairs) in length.

End-to-end aggregation of DNA can be described using models of reversible step-growth polymerization in which polymers of any length can join either end, and a polymer can be broken wherever two monomers are bound end-to-end. Here, a "monomer" is a single 10 bp DNA fragment, and a "polymer" is a chain of end-to-end bound DNA fragments. For analysis purposes, a pair of DNA ends were considered bound if atoms from their terminal base pairs were well-formed and included atoms within 3.75 Å of one another. For a base pair to be considered well-formed we required that the distance between adenine N1 and thymine N3 atoms was 2.7–3.3 Å, averaged over 1 ns. Additionally, we eliminated transient encounters and ruptures by requiring DNA ends to be in contact for at least 2 ns before counting a binding event, and by requiring bound DNA ends to be out of contact for at least 3 ns before counting a rupture. Despite these measures, a handful of false ruptures had to be visually examined before being disregarded for subsequent analysis.

Under the approximation that polymerization and depolymerization rates are independent of chain length, the time-dependent concentration of polymers starting from a solution of monomers can be described analytically [12]. Following the original derivation but under the simplification that no condensate product is produced by polymerization, we obtain solutions for the concentration, c_n, of polymers of length n: $c_n = x(t)y^{n-1}(t)$, with $x(t) = \lambda^2(t)/c_0$, $y(t) = 1 - \lambda(t)/c_0$, initial concentration of monomers, c_0, and the total concentration of polymers, λ. In turn, λ satisfies the 2$^{\text{nd}}$ order rate equation $\frac{d\lambda}{dt} = -k_{\text{on}}\lambda^2 + k_{\text{off}}(c_0 - \lambda)$. At equilibrium, $\lambda = \frac{1}{2K_{\text{eq}}}\left[\sqrt{1 + 4K_{\text{eq}}c_0} - 1\right]$, and, c_n can be expressed as an exponential function that decays with respect to n.

The reversible step-growth polymerization model yields a relation between the mean aggregation number, $\langle N \rangle$, and the equilibrium binding constant, K_{eq}, as a function of DNA concentration, c,

$$\langle N \rangle = \left[\ln \frac{1 + 2K_{\text{eq}}c - (1 + 4K_{\text{eq}}c)^{\frac{1}{2}}}{2K_{\text{eq}}c} \right]^{-1}.$$

In addition to equilibrium properties, the total number of polymers follows second order kinetics and k_{on} can be extracted by fitting $N(t) = \frac{N_0}{1 + tc_0k_{\text{on}}}$ to the data, Fig. S6. Since only one rupture event was observed, k_{off} could not be determined by fitting the data.

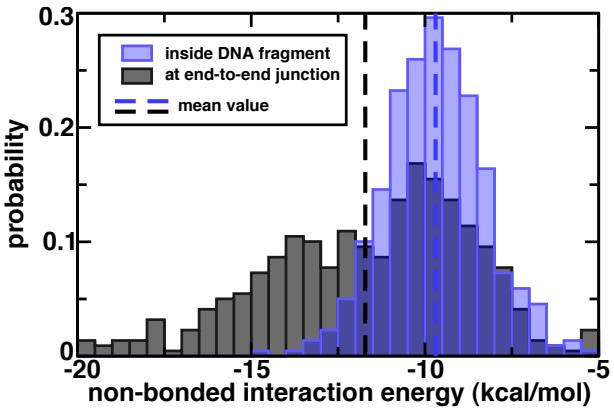

Figure S7: Distribution of the van der Waals and Coulomb interaction energy between two adjacent DNA base pairs forming an end-to-end junction (grey) and buried in the middle of one of the DNA fragments forming an end-to-end assembly (blue). The distributions were obtained from a simulation of two 10-bp DNA fragments forming an end-to-end assembly that was free to translate and tumble.

Modeling rupture as a Poisson process, we found values of k_{off} for which the Poisson distribution $P(r = 1|k_{\text{off}})$ predicts one event with a probability greater than 10%, as described in the main text. While nothing is inherently wrong with this approach, we would prefer to have the likelihood that k_{off} is within some range given 1 rupture event, $P(k_{\text{off}}|r = 1)$. For this, we employed Bayesian inference using the Jeffreys prior and obtained a similar range of values for k_{off}^{-1}, namely 17,000–380,000 ns. We note that the application of Bayesian inference is not well advised given such poor statistics, and that the resulting range for k_{off} depends sensitively on our choice of the prior distribution.

5 Discussion

Figure S7 plots distributions of the non-bonded (Coulomb and van der Waals) interaction energy of adjacent base pairs in the middle of a DNA fragment (blue) and across an end-to-end junction (grey). The distribution was obtained from analysis of a trajectory of a 5′-phosphorylated end-to-end DNA assembly with $\phi = -20°$ (red traces in Fig. 2 d,e of the main text). Only inter-nucleobase interactions were considered; the deoxyribose and nucleic backbone were excluded. The average enthalpy was found to be 2 kcal/mol lower for bases forming the end-to-end junction than for bases in the middle of one of the DNA fragments.

6 Animation descriptions

Animation 1 Animation of a typical simulation trajectory depicting rapid collapse of a pair of DNA fragments into an end-to-end assembly. The 5′-ends of DNA fragments terminated in a hydroxyl group represented as van der Waals spheres.

Animation 2 Animation of a typical simulation trajectory depicting rapid collapse of a pair of DNA fragments into an end-to-end assembly. The 5′-ends of DNA fragments terminated in a phosphate group represented as van der Waals spheres. The negatively charged phosphates guide the DNA into an end-to-end assembly with $\phi \sim -20°$(see Fig. 1 d,e in the main text).

Animation 3 Animation of the simulation that produced the blue trace in Fig. 2 d,e of the main text. The DNA did not have 5′-phosphorylated ends and started with relative azimuthal angle $\phi \sim -20°$.

Animation 4 Animation of an SMD trajectory in which a spring was tethered to the CoM of each DNA fragment. The rest length of the spring was increased, inducing rupture via shearing of the terminal base pairs (see Fig. S1).

Animation 5 Animation of an SMD trajectory in which a spring was tethered to the terminal base pair of on of the DNA fragments. The spring was pulled along the $-x$ axis, causing the end-to-end assembly to shear (see Fig. S2). Phosphorous atoms of the other DNA fragment were harmonically restrained about their initial positions.

Animation 6 Animation depicting a simulation of 458 DNA fragments in a cubic volume 23.8 Å on each side. The DNA fragments, which were not in contact initially, were observed to aggregate forming chains containing up to 11 DNA fragments in length during a 260 ns simulation. Initially, the unit cell containing all the DNA fragments is depicted. Subsequently, the assembly of the longest 10 aggregates is animated. Since aggregates can span across the periodic boundary of the system, neighboring periodic images of the unit cell are shown.

References

[1] Isralewitz, B., Izrailev, S., and Schulten, K. (1997) Binding pathway of retinal to bacterio-opsin: A prediction by molecular dynamics simulations. *Biophys. J.,* **73**, 2972–2979.

[2] Jarzynski, C. (1997) Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.,* **78**, 2690–2693.

[3] Stofer, E., Chipot, C., and Lavery, R. (October, 1999) Free energy calculations of watson-crick base pairing in aqueous solution. *J. Am. Chem. Soc.,* **121**(41), 9503–9508.

[4] Torrie, G. and Valleau, J. (1977) Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics,* **23**(2), 187–199.

[5] Roux, B. (1995) The calculation of the potential of mean force using computer simulations. *Computer Physics Communications,* **91**, 275–282.

[6] Woo, H.-J. and Roux, B. (2005) Calculation of absolute protein–ligand binding free energy from computer simulations. *Proc. Natl. Acad. Sci. USA,* **102**(19), 6825–6830.

[7] Singh, U. C., Brown, F. K., Bash, P. A., and Kollman, P. A. (1987) An approach to the application of free energy perturbation methods using molecular dynamics: applications to the transformations of $CH_3OH \rightarrow CH_3CH_3$, $H_3O^+ \rightarrow NH_4^+$, glycine → alanine, and alanine → phenylalanine in aqueous solution and to $H_3O^+(H_2O)_3 \rightarrow NH_4^+(H_2O)_3$ in the gas phase. *J. Am. Chem. Soc.,* **109**(6), 1607–1614.

[8] Zwanzig, R. (2001) Nonequilibrium statistical mechanics, Oxford University Press, New York.

[9] Wereszczynski, J. and Andricioaei, I. (2010) Free energy calculations reveal rotating-ratchet mechanism for DNA supercoil relaxation by topoisomerase IB and its inhibition. *Biophys. J.,* **99**(3), 869–878.

[10] Cordova, A., Deserno, M., Gelbart, W. M., Gelbart, W. M., and Ben-Shaul, A. (2003) Osmotic shock and the strength of viral capsids. *Biophys. J.,* **85**(1), 70–74.

[11] Hansen, P., Podgornik, R., and Parsegian, V. (2001) Osmotic properties of DNA: Critical evaluation of counterion condensation theory. *,* **64**(2), 1–4.

[12] Kumar, A. and Khanna, A. (1990) Analytical solution of reversible step growth polymerization in semibatch reactors. *J. Appl. Polymer Sci.,* **41**(910), 2077–2094.